(12)

TRAINING EFFECTIVENESS AS A FUNCTION OF TRAINING DEVICE FIDELITY

David R. Baum, Sharon Riedel
Honeywell Systems and Research Center

Robert T. Hays, Angelo Mirabella
Army Research Institute

AD-A140 997

TRAINING AND SIMULATION TECHNICAL AREA

DTIC
ELECTE
MAY 1 1 1984
S   D
B

DTIC FILE COPY

ari

U. S. Army

Research Institute for the Behavioral and Social Sciences

August 1982

84 05 10 009

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

JOSEPH ZEIDNER
Technical Director

L. NEALE COSBY
Colonel, IN
Commander

Research accomplished under contract to
the Department of the Army

Honeywell Systems and Research Center
Minneapolis, Minnesota 55440

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TECHNICAL REPORT 593	2. GOV'T ACCESSION NUMBER Ad-A140 997	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (AND SUBTITLE) TRAINING EFFECTIVENESS AS A FUNCTION OF TRAINING DEVICE FIDELITY		5. TYPE OF REPORT/PERIOD COVERED Final Report March 1981 to August 1982
		6. PERFORMING ORG. REPORT NUMBER 82SRC37
7. AUTHOR(S) David R. Baum, & Sharon L. Riedel (Honeywell Systems & Research Center) Robert T. Hays, & Angelo Mirabella (ARI)		8. CONTRACT OR GRANT NUMBER(S) MDA 903-81-C-0214
9. PERFORMING ORGANIZATIONS NAME/ADDRESS Honeywell Systems and Research Center 2600 Ridgway Parkway, PO Box 312 Minneapolis, Minnesota 55440		10. PROGRAM ELEMENT,PROJECT,TASK AREA & WORK UNIT NUMBERS 2Q162717A790
11. CONTROLLING OFFICE NAME/ADDRESS U.S. Army Research Institute for the Behavioral and and Social Sciences. (PERI-II) 5001 Eisenhower Avenue, Alexandria, VA 22333		12. REPORT DATE 31 August 1982
		13. NUMBER OF PAGES 65
14. MONITORING AGENCY NAME/ADDRESS (IF DIFFERENT FROM CONT. OFF.) ---		15. SECURITY CLASSIFICATION (OF THIS REPORT) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (OF THIS REPORT)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (OF THE ABSTRACT ENTERED IN BLOCK 20, IF DIFFERENT FROM REPORT)

18. SUPPLEMENTARY NOTES

The appendixes to this report have been published as Research Note 82-27

19. KEY WORDS (CONTINUE ON REVERSE SIDE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)

Simulation training	Fidelity specification	Perceptual-motor skill
Simulators	Fidelity requirements	Maintenance training
Training devices	Fidelity research	
Fidelity	Training device effectiveness	

20. ABSTRACT (CONTINUE ON REVERSE SIDE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)

The objective of this study was to determine the effects of reduced training device fidelity on learning and performance of a perceptual-motor maintenance task. Bicycle wheel truing was chosen for study. Five devices including the actual equipment were procured, or designed and built. The device fidelity was systematically varied in physical and functional similarity to the actual equipment. One hundred naive high school and vocational technical school students served as paid subjects; 20 were trained in each device condition. All subjects were then tested on the actual equipment.

(continued)

DD FORM 1473 EDITION OF 1 NOV 55 IS OBSOLETE
1 JAN 73

HD-168 REV 11/74

20. (continued)

The results indicated that significant skill was acquired under all training conditions. The amount of skill acquired did not differ as a function of overall fidelity (i.e., with physical and functional similarity at the same level). However, further analysis in which these two dimensions were separated showed a significant effect of physical similarity. High physical similarity resulted in higher performance on the transfer of training task than low physical similarity.

It was concluded that:

1. The bi-dimensional approach to fidelity is workable at the level of detail required for empirical research.

2. Without an optimized interface and training method, a computer graphics device provides no learning facilitation for this task beyond that found with a set of line drawings.

3. Training a perceptual-motor maintenance task with disabled actual equipment may be as effective as training with fully operational actual equipment.

4. In fidelity research, it is not sufficient to study general levels of fidelity; fidelity must be operationalized in terms of at least two dimensions--physical and functional similarity.

5. Both physical and functional similarity can exist along a number of parameters useful for the purpose of defining training simulator characteristics.

Further research was proposed in the context of specific experiments. Finally, recommendations for the organization and communication of research results via a computerized database were presented.

Technical Report 593

TRAINING EFFECTIVENESS AS A FUNCTION OF TRAINING DEVICE FIDELITY

David R. Baum, Sharon Riedel
Honeywell Systems and Research Center

Robert T. Hays, Angelo Mirabella
Army Research Institute

The Training Technical Area (Simulation Systems Design Team) of the Army Research Institute for the Behavioral and Social Sciences (ARI) performs research and development in areas that include training simulation with applicability to military training. Of special interest is research in the area of simulation fidelity requirements. Before the Army can develop and procure training systems, it must first determine the required level of training system fidelity.

This report provides empirical data on the relationship of simulator fidelity to training effectiveness for a psychomotor task. These data will be entered into a training device research issues data base which will serve as the foundation for user-oriented guidance for the specification of training device characteristics.

User-oriented guidelines for making training simulator design decisions will facilitate the efforts of training device procurers such as the Project Manager for Training Devices (PM TRADE) and also training device users such as the Army Training and Doctrine Command (TRADOC) via the Army's training schools.
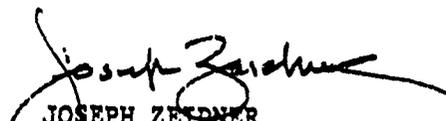
JOSEPH ZEIDNER
Technical Director

Accession For	
NTIS GRA&I	✓
DTIC TAB	☐
Unannounced	☐
Justification	

By_____
Distribution/
Availability Codes

Dist	Avail and/or Special
A-1	

DTIC
COPY
INSPECTED

ACKNOWLEDGEMENTS

TRAINING EFFECTIVENESS AS A FUNCTION OF TRAINING DEVICE FIDELITY

EXECUTIVE SUMMARY

Requirement:

To initiate the development of a database on the relationship between training device fidelity and training effectiveness. To do this by designing and conducting an experiment to explore the effects of reduced training device fidelity on the learning and performance of a perceptual-motor maintenance task. Also, to recommend further research and methods for disseminating research data in the form of guidance to training device developers.

Procedure:

A bi-dimensional definition was adopted to empirically study fidelity. Fidelity was defined as the degree of physical (how it appears) and functional (how it works) similarity between a training device and the equipment being simulated. The perceptual-motor maintenance task chosen for study was bicycle wheel truing. Five devices, including the actual equipment, were procured or specified and built. For three devices, physical and functional similarity were degraded to the same level, that is, high, medium, or low. A computer graphics based device had high functional similarity and low physical similarity, and disabled actual equipment had low functional similarity and high physical similarity.

One hundred subjects were trained. Performance was then tested on the actual equipment and measured as the sum of the peak deviations of the wheel rim from true.

Findings:

Training on all devices led to significant improvements in performance. The results further showed no significant differences in training effectiveness for devices differing in overall fidelity. Thus, the mean performance of subjects trained using line drawings (low physical and low functional similarity) was not significantly worse than subjects trained using the fully operational equipment (high physical and functional similarity). Nonetheless, the performance of the three groups trained on the high, medium, and low fidelity devices was consistently ordered. When the separate effects of physical and functional similarity were analyzed, it was found that the difference between high and low physical similarity was significant; the mean performance of subjects trained on the devices with high physical similarity was higher. High functional similarity did not contribute additional performance benefit.

Utilization of Findings:

These findings can provide an initial entry into ARI's planned computerized database on fidelity and other training system issues. The results may also have immediate implications for the full-scale engineering design phase of the Army Maintenance Training and Evaluation Simulation System (AMTESS) project, even though the current research focused on a limited subset of maintenance behavior. The generalizability of the current research to AMTESS merits analysis. As a corollary, the results have indicated some additional avenues of research on simulator system design issues (e.g., research on applications of computer generated imagery (CGI) to maintenance training and research on media mixes in simulator systems).

TRAINING EFFECTIVENESS AS A FUNCTION OF TRAINING DEVICE FIDELITY

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

PROGRAM OBJECTIVES

In 1981 the US Army Research Institute for t' Behavioral and Social
Sciences (ARI) initiated a research program to improve guidelines for
training device and simulation development. This program, known as
SIMTRAIN, has three major technical objectives:

1.   Evaluate competing methods and models available for use in
      developing and evaluating training devices, and determine
      appropriate applications in the existing acquisition process.

2.   Develop guidelines for relating physical and functional training
      device characteristics (i.e., fidelity) to training effectiveness
      with a focus on maintenance training.

3.   Evaluate the training effectiveness of two alternative versions
      of the Army Maintenance Training and Evaluation Simulation
      System (AMTESS).

This report provides experimental data in support of the second objective.
The remaining objectives are addressed in separate reports.

ARI is pursuing a four-step approach to achieve the second objective--defining
simulator fidelity requirements (Mirabella, 1981):

1.   Abstract principles from existing studies.

2.   Conduct laboratory studies.

3.   Develop a computerized database on fidelity.

4.   Formulate a model and procedure for fidelity analysis.

The end product of this effort will be user-oriented guidelines for generating
fidelity requirements.

1

PURPOSE AND ORGANIZATION OF THIS REPORT

Findings of a laboratory study (step 2) of the relationship between training
device fidelity and training effectiveness are presented in this report. The
present experiment is based on the results of a literature review (step 1) and
a research plan which were documented in a previous report (Baum et al.,
1982). During the development of the research plan, a workshop entitled
Research Issues in the Determination of Simulator Fidelity was conducted. The
proceedings were documented by Hays (1981).

A statement of the problem that gives impetus to the ARI program is presented
in the remainder of this chapter. Background information on the conceptual
framework for this experiment is provided in Chapter 2. The experimental
method and results are described in Chapter 3. In Chapter 4 conclusions and a
re-evaluation of the definition of fidelity are discussed. Finally, in
Chapter 5 a proposal is presented for systematic research on the relationship
between training simulator fidelity and training effectiveness.

PROBLEM STATEMENT

As Baum et al. (1982) state:

> It is widely recognized that simulators and training devices offer
> a potentially cost-effective alternative to training on actual
> equipment. The Army has an increasing commitment to replace or
> supplement hands-on training with training simulators. It is
> therefore necessary, in order to realize the potential increases in
> cost-effectiveness through simulation, to establish a
> systematically and empirically derived database relating training
> simulator configuration and characteristics to training
> effectiveness.

Simulation has a long and accepted (though not uncontroversial) history in
the area of flight training. As the complexity and cost of actual equipment
rises, however, it is becoming increasingly advantageous to apply simulation
approaches to a wider variety of tasks. Equipment maintenance is one such
task domain. With the exception of procedural maintenance tasks, which can
be successfully trained without high fidelity (Baum et al., 1982), very
few data exist to describe the relationship between training device

fidelity and training effectiveness for maintenance tasks. In particular, insufficient research has been conducted on the fidelity requirements for training cognitive (i.e., nonprocedural troubleshooting) or perceptual-motor maintenance tasks.

This research seeks to establish the effect of reduced training simulator fidelity on human performance of a perceptual-motor maintenance task. The task is truing a bicycle wheel. Five different training simulators, including the actual equipment, were specified and built or procured. The five devices represent combinations of different levels of physical and functional similarity to the actual equipment. In the next chapter the conceptualization of fidelity that guided the research effort is described and the rationale for selecting the wheel-truing task is discussed.

CHAPTER 2

BACKGROUND

Two essential components of a research program dealing with the effects of
fidelity on training effectiveness are an operational definition of fidelity
and a task to train. The approach taken to defining fidelity and selecting
a task is discussed in this chapter.

CONCEPTUALIZATION OF FIDELITY

Hays (1980) reviewed the literature on simulator fidelity and showed that a
wide variety of definitions and conceptualizations have been used with the
term. At one extreme the definitions consider the physical similarity of
the simulator to the actual equipment; at the other extreme the definitions
consider the degree to which the trainee perceives the simulator to be a
duplicate of the actual equipment.

Hays proposes that fidelity be limited to descriptions of the simulator and
not be confounded with definitions that incorporate behaviors and
perceptions of the trainee. In a more recent paper (Hays, 1981), he
suggests the following definition of fidelity:

> ...the degree of similarity between the training simulator and
> the equipment which is being simulated. It is a two-dimensional
> measurement of this similarity in terms of:
>
> 1. The physical characteristics of the training simulator
>
> 2. The functional characteristics (i.e., the informational
>   or stimulus and response options) of the simulated
>   equipment

This definition of fidelity (physical and functional similarity) guided the
research effort.

PREVIOUS PAGE
IS BLANK

Physical Similarity

Parameters of physical similarity include size, spatial dimensionality, number and accuracy of details, and accuracy of configuration. All these aspects of physical similarity have been varied in the present study, although they were not varied individually or systematically. Physical similarity is measured here on an ordinal scale and includes three levels—low, medium, and high (Figure 1). Low physical similarity has been operationally defined by a set of line drawings (or computer graphics). Medium similarity is defined by a smaller, degraded version of the actual equipment. High similarity is represented by the actual or target equipment (i.e., the device to which training is being transferred).

Functional Similarity

The functional characteristics of the equipment concern how it works. Functional similarity is defined in terms of the stimulus and response options provided by the device (i.e., how much does it work like the actual equipment). As with physical similarity, functional similarity is also defined on a three-level ordinal scale—low, medium, and high (Figure 1).

Low functional similarity is represented by a simulator that does not work. The trainee's actions on the simulated equipment yield no response—knobs (if there are any) do not turn and buttons do not depress. Medium functional similarity is defined as stimulus options that are available and can be manipulated (the knobs turn, etc.) but do not produce another response from the equipment. High functional similarity is defined as a simulator which provides all stimulus and response options of the actual equipment. The simulator works with effect.

Possible Experimental Conditions

As conceptualized, physical and functional similarity can, in principle, be independently manipulated. It is possible for a simulator to have both low physical similarity and high functional similarity. Low physical similarity

PHYSICAL SIMILARITY

	HIGH	MEDIUM	LOW
	ACTUAL DEVICE	DEGRADED 3-DIMENSIONAL MODEL	PICTURES OR COMPUTER GRAPHICS

HIGH WORKS WITH EFFECT	HH*	HM	HL*
FUNCTIONAL SIMILARITY **MEDIUM** WORKS WITH NO EFFECT	MH	MM*	ML
LOW DOESN'T WORK	LH*	LM	LL*

*DEVICES IN THESE CELLS ARE INCLUDED IN THE PRESENT EXPERIMENT

Figure 1. Possible experimental conditions.

can be achieved in a graphic representation. High functionality can be achieved merely if the trainee has some way to choose a particular stimulus or response option and if choosing the option gives information about the state of the equipment. Thus, high functionality, in terms of the stimulus and response options of equipment, can be provided even in the absence of high physical similarity.

The manipulation of the two dimensions of fidelity is shown in Figure 1. This nine-cell matrix defines a set of devices that could be specified and used in an experiment. Such an experiment would simultaneously provide general information on fidelity (Cells HH, MM, and LL) and specific information on the possibly different effects of physical and functional similarity.

The following are general descriptions of the devices in the nine possible conditions resulting from this conceptualization. Each condition would have a training method and all would be followed oy performance on the target equipment.

Condition HH: High Functional, High Physical Similarity Device--This would be the fully operational, actual equipment or whatever equipment training is being transferred to.

Condition HM: High Functional, Medium Physical Similarity Device--Medium physical similarity is defined as a change in size and the number and accuracy of actual equipment details. A change in dimensionality from three to two is reserved to define low physical similarity. The change in size may be an increase or decrease, but a reduction in size seems intuitively more compatible with the experimental objectives. Likewise, the number and accuracy of simulator details should be less than actual equipment details. Retaining high functional similarity under these conditions requires careful engineering.

Condition HL: High Functional, Low Physical Similarity Device--An effective
means of achieving this condition is through the use of computer graphics and
a software model of the actual equipment. The key change in physical
properties is a change from three to two dimensions. Pictures (line drawings)
are used to represent the actual equipment. High functional similarity is
achieved by allowing the trainee to make stimulus or control choices through a
convenient medium (keyboard, touch panel, etc.). These choices are converted
through a software model of the actual system into appropriate response
information. This information is in turn displayed through the computer
graphics medium; dynamic graphics is used where necessary.

Condition MH: Medium Functional, High Physical Similarity Device--The
requirement here is for the actual equipment to work without effect. This
could most effectively be accomplished by partially disabling the
equipment--disconnecting the displays from the controls but leaving the
control options intact and functional. Knobs, buttons, and other parts would
move but produce no effect in terms of equipment response.

Condition MM: Medium Functional, Medium Physical Similarity
Device--Physically, this device has the same requirements as Condition HM.
The device would be reduced in size and in the number and accuracy of details
compared to the actual or target equipment. Functionally, control choices
could be made, but they would not have any effect on equipment response (see
condition MH).

Condition ML: Medium Functional, Low Physical Similarity Device--The device
in this condition would be a two-dimensional display of the actual equipment
and would provide a means for indicating control choices (e.g., menu selection
through a keyboard input). The simulator would not, however, provide any
equipment response information.

Condition LH: Low Functional, High Physical Similarity Device--This device
would consist of totally disabled actual equipment.

9

Condition LM: Low Functional, Medium Physical Similarity Device--This device would be reduced in size and in the number and accuracy of details compared to the actual equipment, and it would be totally disabled.

Condition LL: Low Functional, Low Physical Similarity Device--This device would consist of a set of line drawings--physically the same as those displayed through computer graphics (see Conditions HL and ML). These line drawings on paper are not functional in any sense.

In Chapter 3 the general device descriptions presented above are implemented in specifications of devices produced or procured for training the wheel-truing task.

TASK SELECTION

A legitimate question is, "Why was wheel truing chosen as the training task?" In this section the task requirements are presented along with a rationale for the selection of the wheel-truing task.

Task Requirements

Baum et al. (1982) discuss the criteria that must be met by any task selected for laboratory experiment in the context of the SIMTRAIN program objectives.

1.   The task must embody the skills required in an actual maintenance task environment.

2.   Task performance must lend itself to straightforward measurement; the measurements must be valid, reliable, and sensitive.

3.   The task must be learnable in a reasonable period of time.

The authors conclude that in order to meet these criteria, it likely will be necessary to study parts of tasks rather than whole tasks.

Rationale for Wheel Truing

Wheel truing is a task that appears to meet the needs of rigorous laboratory research while at the same time requiring performance representative of Army perceptual-motor maintenance tasks.

A description of the wheel-truing task taken from Baum et al. (1982) follows:

> Truing a wheel is not a simple matter. The task is complex enough to be frustrating to a novice, yet appears to be mastered in a reasonable amount of time--after truing 5-10 wheels according to expert opinion.
>
> The task consists of first detecting any misalignment (i.e., correctly attributing wobble to the wheel and not a loose axle), its location(s) and amount, and then manipulating the spoke nipples with a spoke wrench to correct it.
>
> Misalignment is detected by spinning the wheel in the context of fixed reference points on either side of the rim (e.g., the brake pads if the wheel is on the bike). The principle involved in correcting the deviation is to loosen, via the spoke nipples, the spokes that go to the side of the hub that the rim pulls toward and tighten the spokes to the other side. This is a precision operation involving increasingly smaller adjustments of spokes farther away from the point of maximum deviation.

Wheel truing is characterized by the need to adjust and align equipment. The task involves precision eye-hand coordination, a skill component common in perceptual-motor maintenance activities.

In the next chapter an experiment is described that is based on the bi-dimensional conceptualization of fidelity and is carried out in the context of training a wheel-truing task. The experiment is designed to test hypotheses about the relationship between fidelity and training effectiveness.

CHAPTER 3

EXPERIMENTAL METHOD AND RESULTS

The methodology and results of an experiment to examine the effects of reduced training device fidelity on training effectiveness are described in this chapter. The separate and interactive effects of physical and functional similarity were studied in the context of training subjects how to true a bicycle wheel. (See Chapter 2 for a description of the wheel-truing task.) The training effectiveness of five simulators of varying fidelity levels was compared. Subjects were trained on one of the devices and training effectiveness was assessed by comparing their subsequent performance on the actual equipment.

The following null hypotheses were tested:

1. There is no relationship between simulator fidelity and training effectiveness (i.e., the mean performances of the three groups trained on the high, medium, and low (HH, MM, and LL) fidelity devices do not differ).

2. There is no relationship between the physical similarity of the training devices to the actual equipment and training effectiveness (i.e., the mean performance of subjects trained using a high physical similarity device and the mean performance of subjects trained using a low physical similarity device do not differ).

3. There is no relationship between the functional similarity of the training device to the actual equipment and training effectiveness (i.e., the mean performance of subjects trained using a high functional similarity device and the mean performance of subjects trained using a low functional similarity device do not differ).

PREVIOUS PAGE
IS BLANK

METHOD

Approach

The independent variables of physical and functional similarity were
manipulated by constructing simulators that varied systematically along
these dimensions. Five of the nine conditions depicted in Figure 1 were
employed: HH (high functional, high physical), HL (high functional, low
physical), MM (medium functional, medium physical), LH (low functional, high
physical), and LL (low functional, low physical). Subjects were randomly
assigned to one of these conditions, trained to eliminate lateral wobble[*]
in a bicycle wheel using the simulator in that condition, and tested on the
actual equipment to determine the effectiveness of the training.

Subjects

Subjects were obtained through the services of a marketing research firm.
One hundred (85 males and 15 females) non-college bound subjects from
Minneapolis, St. Paul, and the surrounding suburbs were tested. Ages ranged
from 16 to 19 with a mean age of 17.25. There were ten high school
sophomores, 48 juniors, 25 seniors, two high school dropouts, and 15
attending technical school. An additional eight were tested but dropped
from the analysis when it was discovered they were college bound or
previously skilled in bicycle wheel truing. All subjects were tested
individually in one to two hour sessions by a female research associate.

Testing for conditions HH, MM, LH, and LL took place over a four-month
period. Because device HL was not completed until the last month of this
period, all subjects in condition HL were tested in the final month.
Subjects were randomly assigned to conditions with the constraints that each
condition have 17 males, three females, an equal number of technical school
students, and that condition HL subjects be tested in the final month.

---

[*]The complete truing task involves eliminating lateral deviations, making
the rim round, and ensuring that spoke tension is distributed equally.
Thus, in this experiment only a part of the task was trained and tested.

14

Independent Variables

Three independent variables were employed: functional similarity and physical similarity of a training device to the actual equipment, and overall fidelity. Low, medium, and high levels of each variable were used (see Figure 1).

Levels of functional similarity were defined by the degree to which the simulator works with effect. In the low level the simulator does not work; in the medium level the simulator works with no effect; in the high level it works with effect. Low physical similarity is defined by a set of line drawings or computer graphics; the medium level is defined by a smaller stylized version of a bicycle wheel; the high level is defined by the actual physical device. These definitions of physical and functional similarity levels were chosen because they represent general classifications of fidelity that are of practical use to designers of simulators. The intersections of the levels on each variable define the physical and functional characteristics of the simulators. Moreover, the intersections at the high, medium, and low levels of similarity define the third independent variable—overall fidelity.

In this section, each of the devices used in the experiment will be described in detail. General descriptions of devices not used in the experiment but appearing in Figure 1 are found in Chapter 2.

Training Device HH: High Functional, High Physical Similarity—The purpose of the training procedures is to produce skill in truing an actual bicycle wheel. Device HH, the actual bicycle wheel with truing stand, is used both as the training device in condition HH, and as the device to which training is transferred in all conditions. Because device HH is the actual wheel, by definition it represents high physical and high functional similarity.

This device (Figure 2) consists of a bicycle rim with 36 spokes, a truing stand, a spoke wrench, and an electronic mechanism to measure the amount of lateral deviation of the rim from true. The Weinmann rim, size 27 in. by
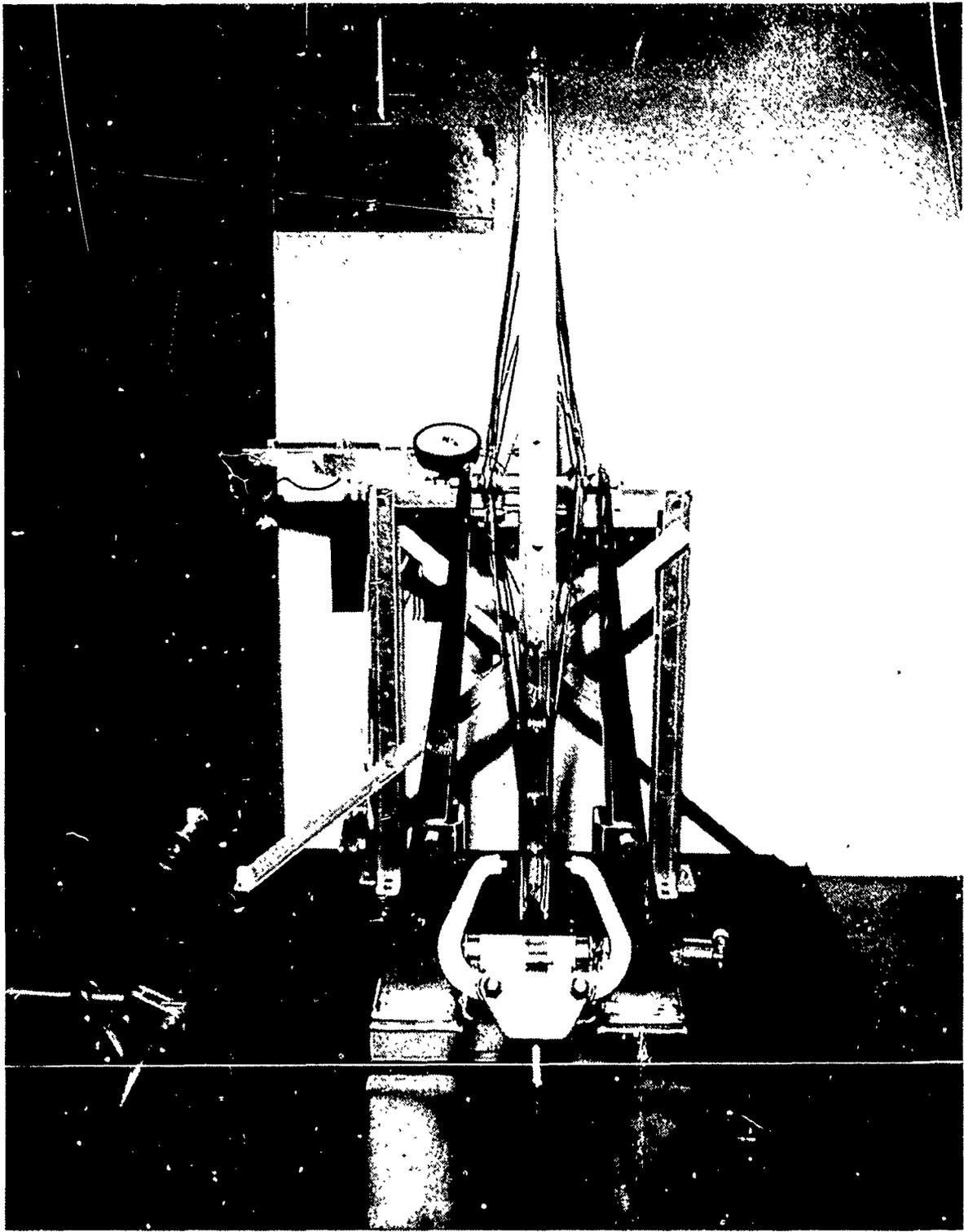
Figure 2. Actual wheel and truing stand.

16

1 1/2 in., has a Normandy hub and solid axle. The truing stand, Park's Model TS2, is an apparatus for holding the wheel vertical while allowing it to spin; an adjustable caliper is used to determine where the rim is out of alignment.

Deviation is measured by the amount of lateral displacement of a 1 in. travel dial indicator (Federal, Model D815) held by a frame against one side of the wheel rim. As the wheel spins, the indicator rod reflects the amount of lateral deviation of the rim. The rod is attached to a linear variable displacement transducer (Schaevitz Engineering, Type 500DC-D C/N 2380), which transforms the lateral movement of the rod to electrical energy. The resulting voltage is fed into a computer and converted into a digital signal. For each measurement the wheel is spun, and the number and magnitude of deviation peaks for one revolution are recorded. When the wheel spins, it passes between an infrared emitting diode and a photo transistor. The spokes block the light between the two devices. This is used to determine when 36 spokes (one revolution) have passed the diode. Specially constructed electronics set the light threshold and remove any false triggering.

Deviation from true is defined as the sum of the absolute values of the deviation peaks for one wheel revolution. Differences less than 0.006 in. were considered to be noise because the rim inherently had that much deviation. The data are stored and available for subsequent statistical treatment. Appendix A documents the user procedures for taking measurements[*]. Appendix B contains a complete listing of the measurement software program.

Training Device HL: High Functional, Low Physical Similarity--Device HL (Figure 3a) is a computer graphics display system that reproduces the line drawings of Device LL (see Figure 5a and 5b) and contains a model of the

---

[*]All appendixes are published together in a separate volume, ARI Research Note 82-27 (AD A133 104).

17

Figure 3a. Graphics display device--subject interface.

18
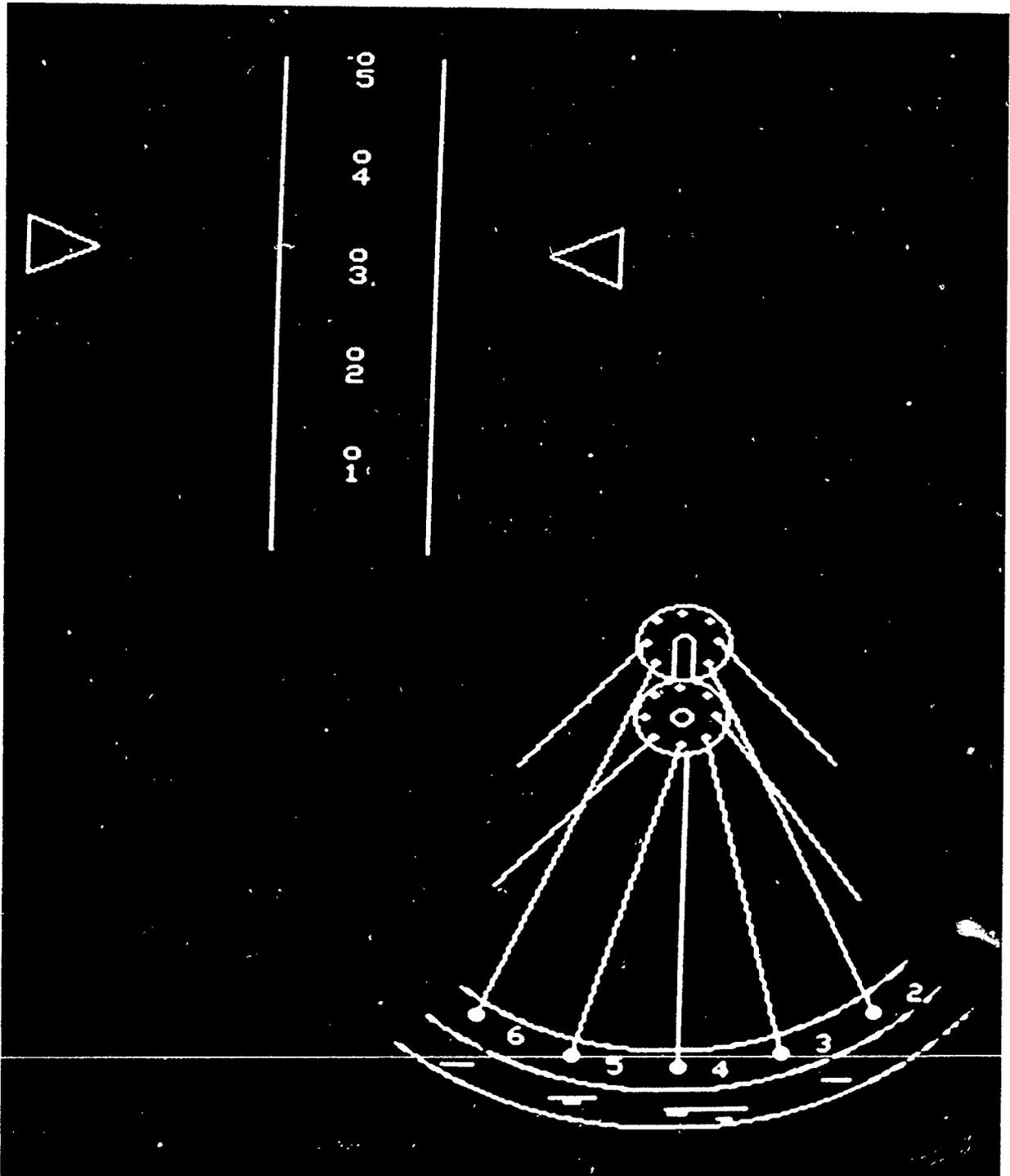
Figure 3b. Graphics display device--dynamic.

wheel rim creating a dynamic version of Figure 5b.* The dynamic graphics (Figure 3b) are adjustable: the wheel can be turned, the caliper can be moved in and out, and the spoke nipples can be turned to effect change in the alignment of the rim in relation to the caliper. The subject, by choosing from a menu of keyboard inputs, is able to make the actual adjustments required to true the rim. The subject can choose to (1) adjust the spokes by turning the nipples clockwise or counterclockwise, (2) adjust the caliper, (3) spin the wheel, (4) stop the wheel, (5) change the direction of the wheel, or (6) change the speed of the wheel. In addition, the experimenter is able to take measurements in a fashion analogous to that on the actual equipment.

Training Device MM: Medium Functional, Medium Physical Similarity--Device MM is a degraded three-dimensional model of a bicycle wheel and stand with a metal rod bent in the shape of a three-sided square representing a spoke wrench (Figure 4). The wheel and stand are constructed of aluminum; the spokes are stainless steel. The wheel is 13 inches in diameter, has eight spokes, and the wheel and stand have a black anodized finish. The nipples are represented by outsize oblong rectangles on the ends of spokes. The nipples can be turned but have no effect on the wheel alignment. The caliper knob turns but does not move the caliper. The wheel turns but the movement can only be used to detect a constant deviation. The wheel and stand are attached to a 10 in. by 20 1/2 in. piece of composition board.

Training Device LH: Low Functional, High Physical Similarity--Device LH is the real wheel and stand as pictured in Figure 2; however, none of the parts move. The caliper connot be adjusted, the wheel is stationary, and the nipples cannot be turned. This is achieved by soldering the spoke nipples and locking the axle when the wheel is placed in the truing stand. The subject is not permitted to turn the caliper knob.

---

*ARI Research Note 82-27 (AD A133 104) contains the appendixes to this technical report, including Appendix C, a complete description of the capabilities of the computer graphics display training device, and Appendix D, a complete program listing for device HL.
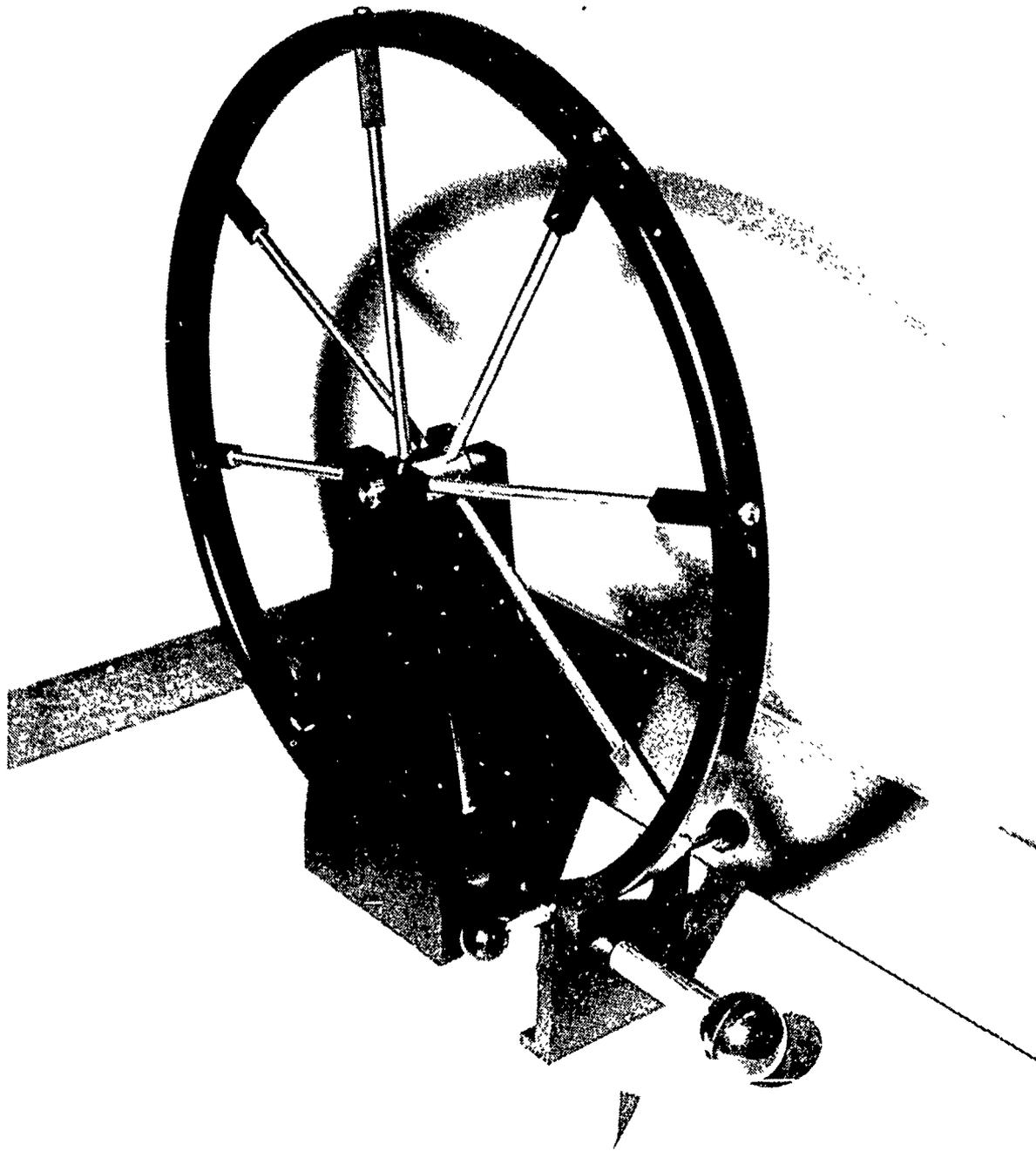
Figure 4. Medium-fidelity device.

21
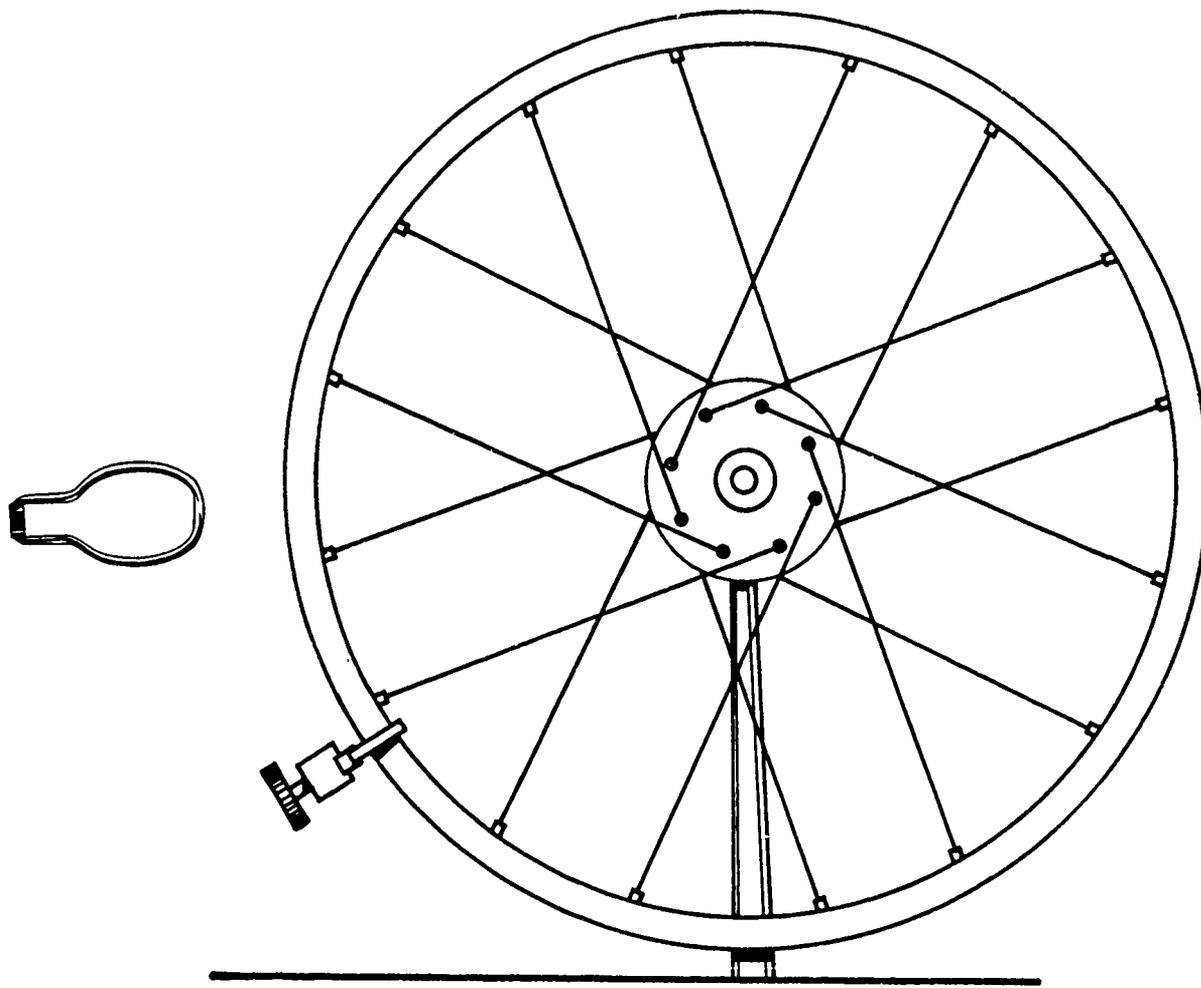
Figure 5a. Low-fidelity device--equipment components.
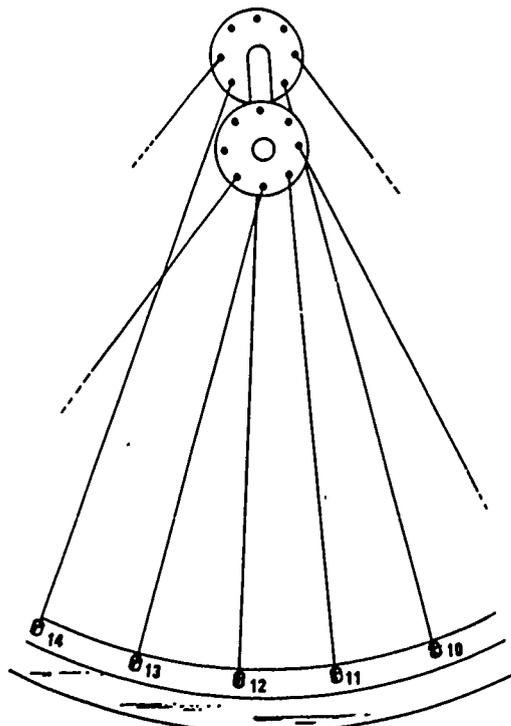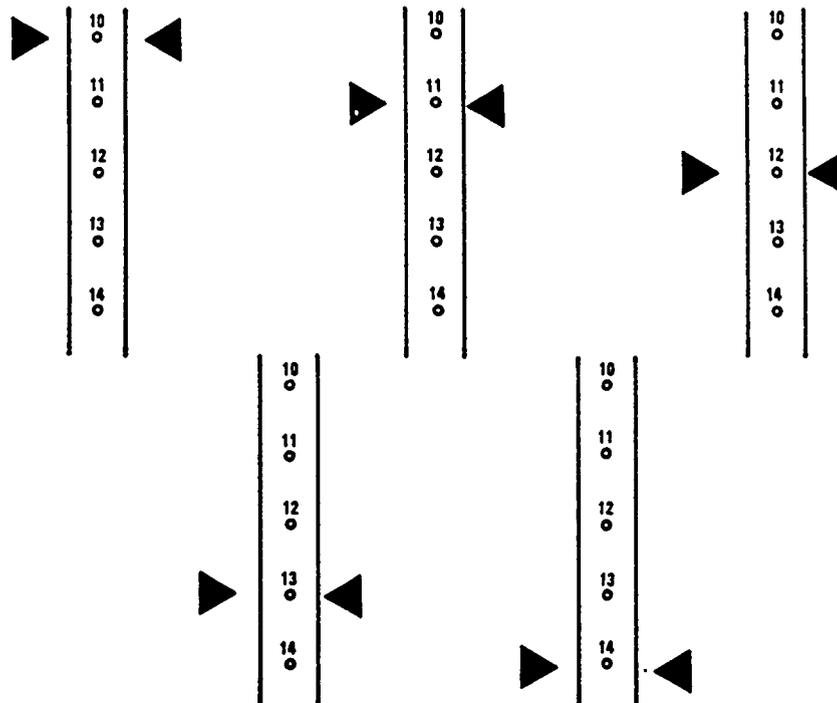
Figure 5b.   Low-fidelity device--simulator exercise.

23

Training Device LL: Low Functional, Low Physical Similarity--Device LL is a set of line drawings as shown in Figures 5a and b and Appendix G. The drawings are two-dimensional representations of parts of the actual equipment. Figure 5a is the first drawing--a picture of the bicycle wheel with parts essential to actual truing. The remaining training illustrations for device LL are included in Appendix G and are designed to facilitate an explanation of how to determine which spokes should be tightened or loosened, how to tighten or loosen the nipples, and how to make differential adjustments of the spokes. The instructions also include a set of five practice exercises for the subject; the first is a demonstration (Figure 5b). Appendix G contains these exercises. They differ only in the position of the caliper.

Figure 6 summarizes the matrix of experimental conditions with a description of the devices used in each.

Objective Assessment of Similarity Levels

The scaling of the independent variables into low, medium, and high similarity was tested by asking four Honeywell training simulator development experts to rate the physical and functional similarity of the five training devices on a scale of one to seven.

Appendix E contains copies of the filled out rating forms. Mean ratings for the five devices are presented in Table 1. These ratings are compatible with and serve as an indpendent validation of the a priori scaling of the devices into low, medium, and high physical and functional similarity.

Dependent Variable

The dependent variable for the experiment was wheel-truing proficiency as measured by the amount of lateral deviation of the wheel rim from true. Before each training and performance trial, the experimenter trued the wheel and introduced a standard amount of deviation. Deviation was measured at the start of each training and performance trial and at three-minute

PHYSICAL SIMILARITY

	HIGH ACTUAL DEVICE	MEDIUM DEGRADED 3- DIMENSIONAL MODEL	LOW PICTURES OR COMPUTER GRAPHICS
HIGH **WORKS WITH** **EFFECT**	**HH** ACTUAL BICYCLE WHEEL WITH TRUING STAND		**HL** COMPUTER GRAPHICS DISPLAY WITH A DYNAMIC MODEL OF THE WHEEL RIM. WORKS LIKE THE ACTUAL EQUIPMENT.
MEDIUM **WORKS WITH** **NO EFFECT**	.	**MM** SMALLER, SIMPLER MODEL OF BICYCLE WHEEL AND TRUING STAND. PARTS ARE MOVABLE BUT HAVE NO EFFECT ON ALIGNMENT.	
LOW **DOES NOT** **WORK**	**LH** TOTALLY DISABLED ACTUAL BICYCLE WHEEL AND TRUING STAND		**LL** SET OF LINE DRAWINGS OF THE WHEEL AND STAND THAT SHOWS PARTS AND FACILITATES AN EXPLANATION OF TRUING.

FUNCTIONAL SIMILARITY (vertical axis label)

Figure 6. Devices used in the experiment.

25

TABLE 1. MEAN RATINGS* OF PHYSICAL AND FUNCTIONAL
SIMILARITY OF TRAINING DEVICES

Device	Physical Similarity Mean	Functional Similarity Mean
HH (High Functional, High Physical)	7	7
HL (High Functional, Low Physical)	2.25	5
MM (Medium Functional, Medium Physical)	3.5	3.5
LH (Low Functional, High Physical)	5.5	1.75
LL (Low Functional, Low Physical)	2	I.75

*Raw data are included in Appendix E.
(See Research Note 82-27)

intervals auring the 15-minute trials. For each measurement, the sum of the absolute values of peak deviations in thousandths of an inch was computed and stored. Values under 0.006 in. were treated as noise and did not enter into the sums.

Procedure

Subjects were randomly assigned to one of five training devices. Training on these five devices made up the five experimental conditions (see Figures 1 and 6). Each experimental condition consisted of two parts--a training phase and a performance phase (see Table 2). The training phase consisted of a brief experimenter demonstration of wheel truing and a practice session on the appropriate device. (Appendix H presents a description of the training instruction.)

In the demonstration, the relevant parts of the wheel and truing equipment were pointed out and the wheel truing procedure described. The subjects then used the device in their experimental condition to practice wheel

truing. Practice time in conditions MM, LH, and LL was approximately 5
minutes; conditions HH and HL practice time was 15 minutes. Subjects in
all conditions were allowed to ask for assistance during the practice
sessions.

Subjects in condition HH practiced on the actual wheel. They were given one
15-minute practice trial.[*] Before each practice trial for each subject,
the experimenter trued the wheel and introduced a standard amount of
deviation. Deviation measurements were taken prior to each trial and at
three-minute intervals during the practice trial of conditions HH and HL,
but they were not analyzed. Practice measurements could not be taken in the
other conditions (MM, LH, and LL).

All subjects were then given two 15-minute performance trials on the actual
equipment. Subjects were not permitted to ask for assistance during the
performance session. Before each performance trial for each subject, the
experimenter trued the wheel and introduced a standard amount of deviation.
Measurements of wheel deviation were taken before each trial and at three
minute intervals during the trial. Table 2 summarizes the experimental
procedure.

To establish a performance ceiling, data were also collected on three wheel
truing experts who performed two 15-minute trials. These data were
subsequently used to rule out an alternative explanation for a finding of no
differences among groups, i.e., the results might be attributable to the
groups reaching a point where no further improvement was possible.

Treatment of the Data and Statistical Analyses

The data analyzed were the sums of the absolute values of deviation peaks
for one revolution of the wheel taken initially and at three-minute

---

[*]Pilot research had shown that 15 minutes was necessary for performance to
asymptote. All practice and performance trials on Device HH were therefore
fixed at 15 minutes duration.

TABLE 2. EXPERIMENTAL PROCEDURE

| Experimental Conditions | Training | | Performance Trials All on Device HH (Actual Equipment) | |
| | Demonstration | Practice on Device | 1 | 2 |
			Measurements	Measurements
HH High Functional High Physical Similarity (N = 20)		15-minute exercise	15 minutes	15 minutes
LL Low Functional Low Physical Similarity (N = 20)		5-minute exercise		
HL High Functional Low Physical Similarity (N = 20)		15-minute exercise		
MM Medium Functional Medium Physical Similarity (N = 20)		5-minute exercise		
LH Low Functional High Physical Similarity (N = 20)		5-minute exercise		

intervals during each trial for each subject. The two performance trials resulted in 12 such measurements for all subjects. The performance trial 2 data of one subject were lost due to a computer failure.

The data were first examined to determine if subjects showed significant improvement in performance during each performance trial. Ten t-tests were computed comparing initial setting with final measurement. This was done for performance trials 1 and 2 for each of the five conditions.

Two ANOVAS were conducted. In the first, overall fidelity served as the non-repeated independent variable. In the second, the non-repeated variables were functional and physical similarity. The repeated measure in both ANOVAS was trial. The final measurements from performance trials 1 and 2 provided the data analyzed.

Because the final measurement of each trial was selected as the primary dependent variable, it was important to determine if this measurement was influenced by the initial (measurement) setting. Pearson r correlations were computed between the initial and final measurements of each trial. For performance trial 1, $r = 0.020$ ($N = 100$); for performance trial 2, $r = 0.018$ ($N = 99$). These correlations are not statistically significant, and any final measurement differences among groups cannot be attributed to group differences on initial measurements.

Because device HL (the computer graphics device) was not completed until the last month of the four-month testing period, all condition HL subjects were tested in this final month. To determine if performance differences between condition HL subjects and the other groups might be due to differences between the subjects tested in the first and second time periods, the performances of subjects in these two time periods were compared. Subjects in conditions HH, MM, LH, and LL were pooled and then divided into two groups--those tested during the first three months ($n = 45$) and those tested during the final month ($n = 35$). The early and late groups were compared on the final measurement of performance trials 1 and 2. Unweighted means F tests (Keppel, 1973) show no differences between the performances of the

29

early and late subjects. For trial 1, $F(1, 78) = 2.65$, $p > 0.05$; for trial 2, $F(1, 78) = 3.36$, $p > 0.05$. Any differences between the groups cannot then be attributed to time of testing.

RESULTS

As can be seen in Table 3, all of the t-tests comparing the initial and final measurements of each group were statistically significant at $p < 0.005$. The significance levels have not been adjusted to reflect the computation of multiple t-tests. However, any adjustment made will still yield significance levels of at least 0.05. These results indicate that regardless of the training device used, subjects' mean performance improved significantly over the course of each trial. All of the devices were therefore effective in training-wheel truing.

Next, the data were analyzed to examine the relationship between fidelity and training effectiveness. A oneway repeated measures ANOVA (BMDP2V) was performed using conditions HH—high, MM—medium, and LL—low, as the between groups factor and trials 1 and 2 as the repeated measures. As can be seen

TABLE 3. COMPARISONS OF INITIAL AND FINAL MEASUREMENTS
FOR TRIALS 1 AND 2 BY CONDITION

| Condition	Performance Trial 1			Performance Trial 2		
	n	t	p^*	n	t	p^*
1	20	9.34	<0.001	20	5.22	<0.001
2	20	5.81	<0.001	19	4.13	<0.001
3	20	3.67	<0.001	20	6.41	<0.001
4	20	10.07	<0.001	20	15.96	<0.001
5	20	3.14	<0.003	20	6.81	<0.001

*For a one-tailed t-test.

from Table 4, null hypothesis 1 was not rejected--the main effect of fidelity was not statistically significant. The general level of fidelity does not appear to affect performance.

Figures 7a and 7b present the mean rim deviations for performance trials 1 and 2 for the expert group, and conditions HH, MM, and LL. Although differences between the three conditions are not statistically signficicant, it should be noted that the groups' performances show a consistent ordering from low to high throughout virtually all of the measurements on performance trial 1. The HH condition retains this consistent superiority in performance trial 2. The performance of the expert group shows that the lack of a significant fidelity effect is not due to a ceiling effect--none of the other groups performed as well as the expert group.

Finally, the effect of physical and functional similarity of the device on training effectiveness was examined. A two-way repeated measures ANOVA (BMDP2V) was run on the final measurement of each performance trial, using physical and functional similarity as the between groups variables and performance trials 1 and 2 as the repeated measure. As can be seen from

TABLE 4. REPEATED MEASURES ANOVA ON FINAL MEASUREMENT FOR CONDITIONS HH, MM, AND LL (performance trials 1 and 2)

Source	df	MS	F	P
Condition	2	0.236	1.108	NS
Error	57	0.215		
Trial	1	0.156	2.108	NS
Trial x Condition	2	0.0052	0.703	NS
Error	57	0.0074		
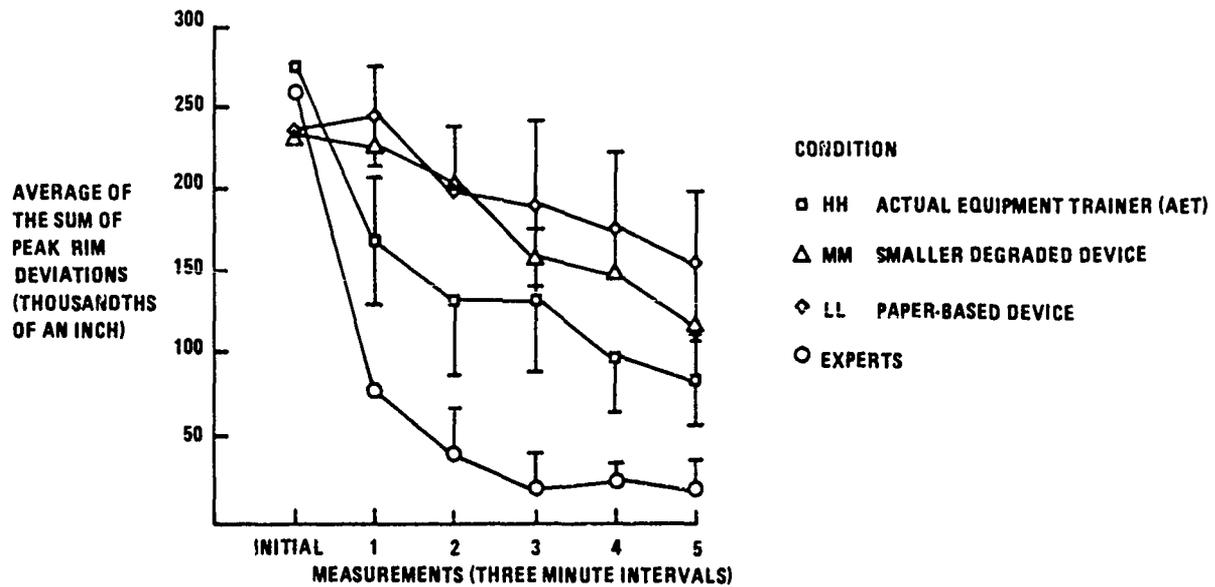
Figure 7a. Mean rim deviations by condition for performance trial 1 (95% standard errors shown for selected means).
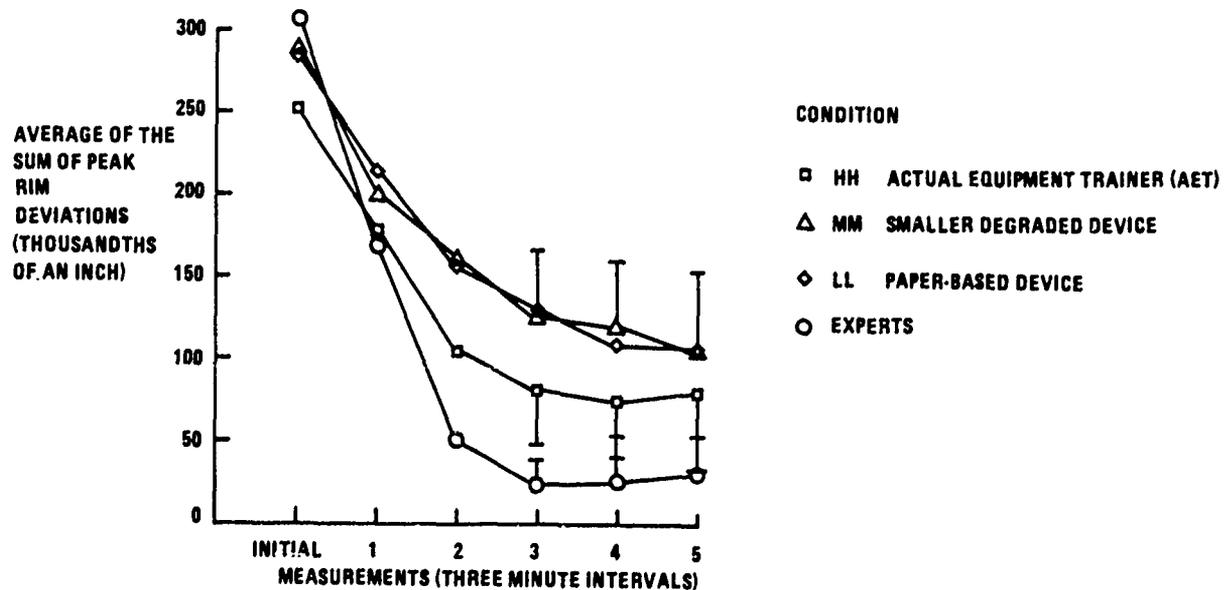
Figure 7b. Mean line deviations by condition for performance trial 2 (95% standard errors shown for selected means).

Table 5, null hypothesis 2 was rejected. The main effect of physical similarity is statistically significant, $F(1, 75) = 4.157$, $p < .05$. Neither the main effect of functional similarity nor any of the interaction effects are significant; thus null hypothesis 3 cannot be rejected. This indicates that for training of this task, effectiveness is a function of the physical similarity of the training device to the actual equipment, but is not affected by functional similarity.

Figure 8 shows the mean rim deviations by condition averaged over performance trials 1 and 2. The differential effect of high and low physical similarity can be clearly seen here. Subjects trained on the two low physical similarity devices performed worse than subjects trained on the two high physical similarity devices. Low and high functional similarity subjects performed equally well.

TABLE 5. REPEATED MEASURES ANOVA ON FINAL MEASUREMENT FOR CONDITIONS HH, HL, LH, AND LL (Performance trials 1 and 2)

Source	df	MS	F	P
Functional Similarity	1	0.006	0.032	
Physical Similarity	1	0.0769	4.157	<0.05
Physical x Functional	1	0.0005	0.027	
Error	75	0.0185		
Trial	1	0.0118	2.034	
Trial x Functional	1	0.0136	2.345	
Trial x Physical	1	0.0003	0.052	
Trial x Functional x Physical	1	0.0043	0.741	
Error	75	0.0058		

CONDITION

□ HH ACTUAL EQUIPMENT TRAINER (AET)

■ LH NONOPERATIONAL AET

◆ HL COMPUTER GRAPHICS DEVICE

◇ LL PAPER-BASED DEVICE
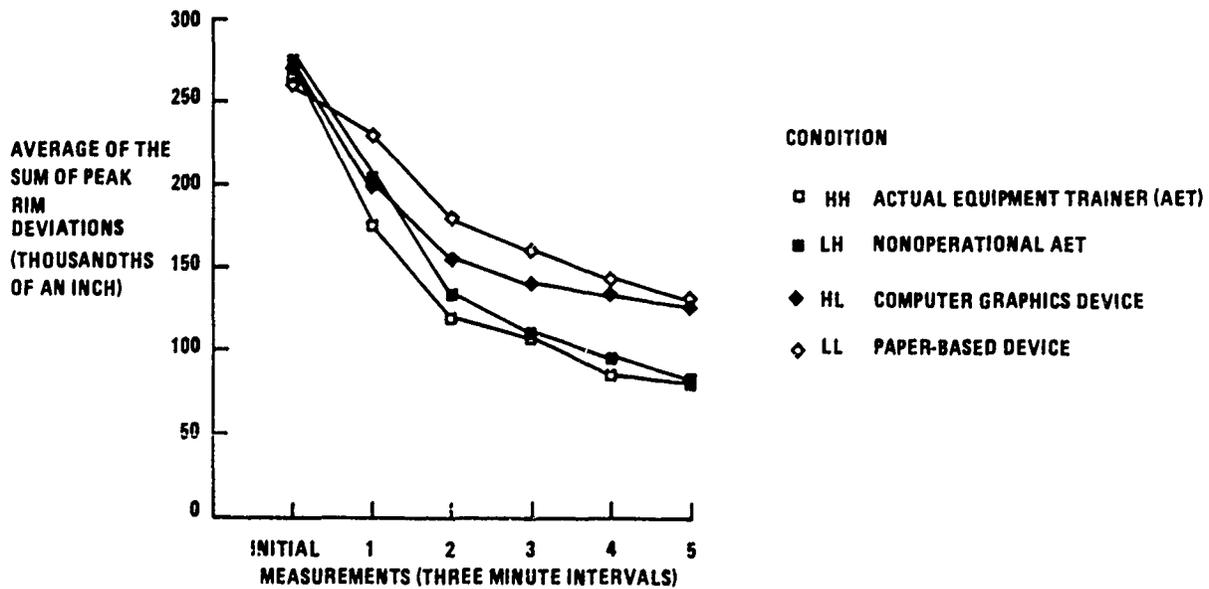
Figure 8. Mean rim deviations by condition, averaged for trials 1 and 2.

34

CHAPTER 4

DISCUSSION AND CONCLUSIONS

The experiment and its results have implications for continued research on training device fidelity. First, the feasibility of studying the separate effects of physical and functional device similarity has been demonstrated. Second, the specific results provide tentat've guidance for training device developers concerned with perceptual-motor maintenance tasks. Third, the research definition of fidelity needs to be re-examined.

FEASIBILITY OF FIDELITY RESEARCH

The general approach taken to conceptualize fidelity proved to be workable at the level of detail required for empirical research. Training devices varying systematically in degree of physical and functional similarity to the actual equipment were procured, or designed and built. The a priori classification of the resulting devices was validated through independent observer judgment.

It is important to note that there was no attempt to compensate for the limitations of each device through different training approaches. Rather, the same general training method was employed throughout. It was hoped that the effects of reduced fidelity would be directly attributable to device characteristics.

In this study the interface between trainee and device was not optimized. Such optimization cou'd involve either the manner of interaction with the device or the degree of training assistance provided by the device. For example, in the first case, instead of a menu and keyboard, interaction with the computer graphics device could be based on a touch panel. This would reduce or eliminate the amount of learning needed to indicate response options. It is possible that subjects in the present experiment were trying

to learn two tasks at once and consequently did not learn either one well. An optimized interactive interface that is easy to use would minimize this possibility.

The second case of optimization would involve use of the computer to monitor performance and deliver feedback. In principle, this should facilitate learning because the computer can detect every error that is made and correct it immediately. Immediate feedback such as this has been shown to be important, especially in initial skill acquisition. Using the computer to provide feedback also relieves the instructor (experimenter) from having to closely monitor performance of a single trainee (subject).

Based upon the present research, it is too early to discount the possible benefits of a computer-based device for training a perceptual-motor maintenance task. One result is clear--without an optimized interface and training method, the computer graphics device provides no learning facilitation for this task beyond that found with a set of line drawings.

These are matters for further research. This research is particularly important in the context of the US Army's program to develop a computer-based generic maintenance training and evaluation simulator system (AMTESS). Issues regarding the optimization of interface design need to be carefully considered before committing a device like AMTESS to full-scale engineering development.

LEVEL OF FIDELITY FOR TRAINING TO PERCEPTUAL-MOTOR MAINTENANCE TASK

This experiment dealt with both overall device fidelity (high, medium, and low) and with the potentially separate effects of two fidelity dimensions-- physical and functional similarity. Training on all five devices led to significant improvement during performance trials. Even subjects trained using line drawings of the equipment showed transfer of learning.

Despite the consistent ordering of performance of subjects trained on devices HH, MM, and LL, the general effect of fidelity did not achieve statistical significance. As with any null effect, we must cautiously interpret its meaning. We cannot state that training on devices of different levels of fidelity is equivalent for this task; we can only state that the differences are not statistically significant. Even the smallest numerical difference can be shown to be statistically significant with a large sample size. In this study, a sample size of approximately 50 would have been sufficient.

What is more important is the practical significance of differences. In this study we wish to generalize to the population of Army maintenance technician trainees. The Army trains thousands; thus, even small differences in effectiveness among training devices might have practical significance (e.g., in terms of cost or readiness). At this stage it is difficult to extrapolate to an impact on maintenance activities in a modern military force. The consistent ordering among groups, although statistically nonsignificant with our sample size, reminds us that we must be very cautious about the lack of a general fidelity effect.

While the general level of fidelity had no effect on training effectiveness, with fidelity divided into two dimensions it was shown that physical similarity has a significant impact on training effectiveness; functional similarity has no effect. This seeming paradox, that general fidelity does not achieve significance while one of its dimensions does, is attributable to the high variability within each group and the increased degrees of freedom and estimation precision that comes from combining groups to assess the effects of physical and functional similarity.

The learning benefit derived from high physical similarity is persistent; it is present at the end of the second practice trial. Thus the effects of different training devices are not eliminated by the interpolated experience on the actual equipment (the first performance trial). The failure of the dimension of functional similarity to reach significance means that training

a perceptual motor maintenance task with disabled actual equipment may be potentially as effective as training with functional actual equipment.

Previous research on the effects of device fidelity has generally resulted in no significant differences in training effectiveness between actual equipment and device-trained subjects (c.f., Orlansky and String, 1981). In this experiment we have demonstrated that when fidelity is partitioned into physical and functional dimensions, significant differences emerge on one dimension. This finding has broad implications for the conduct of future research on the relationship between device fidelity and training effectiveness. It is not sufficient to study general levels of fidelity. Fidelity must be operationalized as consisting of at least two dimensions—physical and functional similarity.

DEFINITION OF FIDELITY

During this research, decisions were made regarding how to implement level of device fidelity. The conceptualization of fidelity, as consisting of device physical and functional similarity to actual equipment, is general. For each device, detailed specifications had to be prepared. Perhaps the clearest issue to emerge from this process is the level of detail to incorporate into a two-dimensional (pictorial) representation of actual equipment.

Physical similarity can exist along a number of parameters. Among these are size, spatial dimensionality, number and accuracy of details, and accuracy of configuration. A device may have low similarity on one of these parameters (e.g., two-dimensional as opposed to three-dimensional) yet have high similarity on another (e.g., the two-dimensional representation may be a photograph).

For practical purposes, one of these parameters may be more important than the others. Perhaps the importance of a parameter of physical similarity depends upon the particular task being trained. For example, for a perceptual-motor task, the results of this experiment would predict that a

photographic representation (high similarity on number and accuracy of details) would yield better performance than the line drawings of device LL. Furthermore, we would predict that using video pictures (analogous to photographs) with a high functional interface would result in no additional performance advantage. However, because a performance benefit will accrue with video-based devices (recall that all groups showed significant learning) if there are non-learning advantages to this kind of device, then it might warrant implementation. For example, if handling a large student flow and enhancing motivation were important, and if skill mastery was not required, then the video device might prove cost effective.

Although not a direct outgrowth of this experiment, another definitional issue concerns the parameters of functional similarity. The issue arises in the context of training cognitive compared to procedural or perceptual-motor tasks, general skills (e.g., troubleshooting) compared to system-specific skills, and experts compared to novices. For these purposes, a kind of functional similarity, more related to wiring diagrams than to front-panel layout and more related to functional interrelationships than to stimulus-response options, seems useful. We might distinguish between concrete and abstract functional similarity. This distinction would correspond to that between informational and stimulus-response options in Hays' (1981) terminology.

Several training devices have been built that are designed to achieve high abstract functional similarity. These derive from the hypothesis that principles of equipment operation or troubleshooting can be most efficiently taught in the context of system models. Two examples STEAMER and FAULT (Framework for Aiding and Understanding Logic Troubleshooting) will be discussed.

STEAMER is a system being built by the Navy and is designed to train students in the principles of propulsion engineering (Williams, et al., 1981). It is based on a mathematical model of an existing full-scale, mock-up simulator of a 1,200-psi steam plant. The model is interfaced to the trainee through computer graphics which present a "wiring" diagram of

the system components. The trainee can manipulate the simulated steam plant by opening or shutting valves and turning components on or off. This is done through a touch-screen interface to a command menu.

A second example comes from the work of Rouse and his colleagues (Rouse, 1979; Hunt and Rouse, 1981; and Johnson and Rouse, 1980). In their research the trainee is presented with fault diagnosis problems via a computer graphics display. Early research was conducted with a context-free display based on computer-wiring diagrams. FAULT, developed later, is a general computer program that can be used to represent malfunction data for various types of engines. Through a keyboard interface, the trainee can gather information about the malfunction, act on the information, and receive feedback about the results or costs of the action.

As training devices, STEAMER and FAULT have low physical similarity and high functional similarity to the actual equipment. Thus, they are like device HL (the computer graphics device). However, the similarity extends beyond the concrete functional dimension, to similarity in the inner workings and dynamics of the propulsion plant and engines.

Guidelines for specification of training device characteristics should allow this aspect of functional similarity to be considered. However, the empirical research on the utility of high abstract functional similarity is inconclusive (Johnson, 1980). At the present time, further research is warranted and not the development of specific guidelines.

CHAPTER 5

PROPOSAL FOR FUTURE RESEARCH

The present experiment is a small but significant step in the development of an empirical database that can aid decisions about training device characteristics. More research is required, however. The database must be systematically expanded; therefore, the necessary research must be well planned. The purpose of this chapter is to present a proposal for a systematic program of research based in part on the framework discussed in Chapter 2. The research proposal, in addition, has been influenced by the results of the present experiment and by the proceedings of the workshop reported by Hays (1981). Above all, the specific independent variables recommended for study are directly tied to key issue areas in army training.

RESEARCH PROPOSAL

This section is divided into four parts. The approach taken to defining the recommended program of research is discussed in the first part. In the second part categories of possible independent variables are presented. Specific experiments are proposed in the third part. Finally, the fourth part contains a discussion of how the results of this and other studies can be organized and disseminated to guide future training device design decisions.

General Framework

This experiment was concerned specifically with the effects of reduced training device fidelity on training effectiveness for a perceptual-motor maintenance task. A key aspect of this experiment was the independent manipulation of physical and functional similarity and the investigation of their separate effects. This will be an important component of the proposed research. Future research efforts must also examine other task domains and

task types. In addition, the effects of additional independent variables within a particular task domain or type must be assessed.

The framework we propose is based on three task dimensions: task domain, performance context, and task type (Figure 9). There are basically three task domains in the military environment where personnel interface with systems: operation, maintenance, and command and control (c^2). The c^2 task domain is characterized by the use of information compared to the operation role of creating, seeking, or gathering information. Each of these domains can involve the performance of an individual or CGTU (crew, group, team, unit). Finally, as discussed in Baum, et al. (1982), the three generic task types of primary interest to military job performance are procedural, perceptual-motor, and cognitive.

Figure 9 indicates these three dimensions may be viewed as independent of one another. The present experiment, for example, falls into the maintenance, individual, perceptual-motor cell. In principle, each of the 18 cells in the matrix provides a candidate set of issues for a fidelity
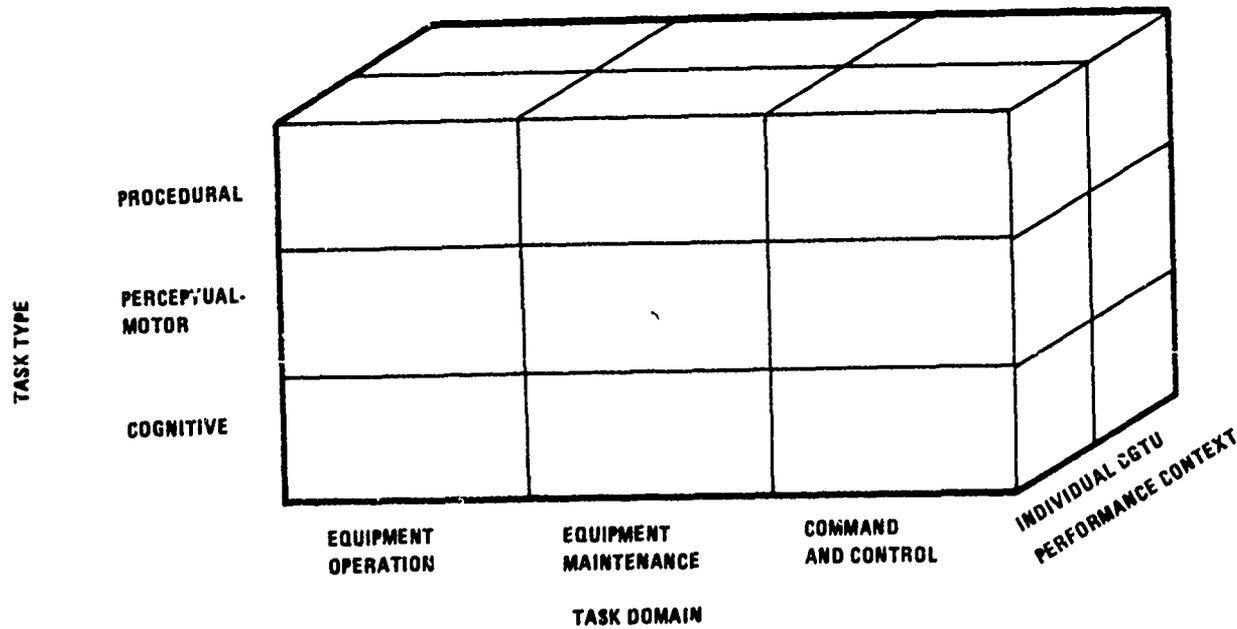

Figure 9. General framework for fidelity research plan.

42

research program. In practice, however, the categories on the dimensions tend to be correlated. For instance, C^2 tends to involve more cognitive CGTU performance than individual procedural tasks. Correlations such as these help set research priorities.

This framework is presented not with the idea of developing an all encompassing research proposal but as a means of organizing existing data and indicating where additional data is needed. The effects of fidelity on training device effectiveness, and thus fidelity requirements, may be quite different depending on the particular cell one is concerned with.

The proposal developed in this chapter deals exclusively with the equipment maintenance domain and individually performed perceptual-motor and cognitive tasks. Before discussing the specific suggested experiments, we will discuss candidate independent variables.

Candidate Independent Variables

The general research question is whether fidelity interacts with other training environment variables in its effect on training effectiveness. The additional independent variables and how they are manipulated (e.g., operationalization, selection of levels, etc.) must reflect Army training problems and priorities.

Perhaps the most significant problem faced by today's Army is the relatively low intelligence and learning ability of its recruits. For example, in FY1981, 34% of all Army recruits were in AFQT Category IV, below average in trainability (Office of the Assistant Secretary of Defense, 1982). A low re-enlistment rate exacerbates this problem by depleting the force of trained personnel. The need is to achieve effective training in the shortest possible time.

Another significant problem is the increasing complexity of the equipment being deployed. In a future conflict, sophisticated technology will balance the scale against large numbers only if soldiers are able to operate and

maintain complex weapon systems. Although our technology should make the use and maintenance of these systems easier, this has not been the case thus far. Despite intensive attempts to adapt procedures to troubleshooting tasks and create step-by-step job aids, many failures still require a technician to make decisions about system repair.

From these problems emerges the need to examine the (training) effects of fidelity in the context of additional independent variables. The categories of these variables include principally trainee characteristics and task difficulty.

Trainee Characteristics--There are three characteristics of the learner that are of concern: general intelligence, aptitude, and level of skill. It is possible, if not probable, that each of these factors will interact with fidelity and may have different effects for physical and functional similarity. For example, it seems plausible that highly intelligent individuals would benefit equally from training devices of different levels of fidelity but low intelligence individuals would not. The Army trains individuals with a wide range of intelligence, aptitude, and skill, albeit in disproportionate numbers. Thus, guidance for device development must accommodate the extremes in trainee characteristics.

Task Difficulty--The two factors of primary importance in the difficulty of the task to-be-trained are complexity and environment. The complexity of the task itself is governed by criteria such as the amount of information (number of alternatives), and the imposition of a time limit or accuracy criterion. Adverse conditions in the environment can also influence task difficulty. For example, as illumination departs from adequate levels, performance suffers; as temperature and humidity increase, physical and mental performance become more difficult (c.f., McCormick, 1970).

Variables of both complexity and environment need to be explored in the context of fidelity manipulations. Intuitively, it seems likely that simple and complex tasks will differ in the degree of fidelity required for

44

effective training. Environmental variables may be more interesting to manipulate during performance (transfer) than during training. In this manner, the resistance of learning to stress can be studied.

Specific Suggested Experiments

The experiments suggested in this section fall into three categories according to the task to be studied. The first category consists of experiments which utilize a more complex wheel-truing task. The second category involves another perceptual-motor maintenance task. The final category is concerned with cognitive maintenance tasks.

Further Experiments: Wheel-Truing--As noted elsewhere, the present experiment might be made more sensitive to training device effects by increasing the complexity of the task. Therefore, the two experiments outlined below would utilize wheel-truing, and the dimensions of alignment (i.e., the dimension used in the present experiment) and roundness would be trained and measured. Technically this is a straightforward extension of the current methodology.

Experiment 1--The first experiment will partially replicate the present results with a more complex version of the wheel-truing task. Only four devices would be utilized: HH, HL, LH, and LL (see Figure 6 and Chapter 3). The rationale for excluding device MM is that at this stage the effects of the extremes are of primary interest.

The subject's task would be nearly the same--only measuring the roundness of the wheel would be added. The training method would be modified accordingly. This would require changing devices HL and LL to add graphics and line drawings that depict departures from a round rim. The subject would be trained to eliminate both kinds of deviation.

It is estimated that between 20-25 subjects per condition would be sufficient. They would be selected according to the same criteria used in this experiment. The training would be conducted as described in

Appendix H. Training on each device would be followed by two fifteen-minute performance trials. Expert performance data would be collected as before.

The analysis would be conducted in the same manner as the present experiment. Three-way (physical similarity x functional similarity x trial) analyses of variance would be conducted for the final measurement on both dependent measures (alignment and roundness).

This experiment is expected to result in a more reliable effect of physical similarity. Also, the effect should be strong enough to show a difference between devices HH and LL.

Experiment 2--The objective of this experiment will be to determine if the effects of fidelity are different for subjects who differ in general intelligence and aptitude. The design described for Experiment 1 would be employed. Subjects, however, would be selected according to scores achieved on the Armed Services Vocational Aptitude Battery (ASVAB). Four groups of subjects would be formed based on trainability category (Cat 1 + 2 vs Cat 4) and mechanical aptitude* (hi vs low). Each group would consist of 80 individuals (320 total) and 20 from each group would be trained on each device.

Subjects for tnis experiment would ideally come from the armed forces recruit population. Test scores would already be on file for this population. Authorization and close contact with the Recruiting Command would be required to procure recruits before they went on active duty status. Adequate numbers should be available through various delayed or deferred enlistment programs. The drawback of this approach is the need for agreement and close coordination among agencies with diverse requirements and restrictions.

---

*Mechanical Aptitude is a composite of Mechanical Comprehension, Automotive-shop Information, and General Science subtests.

If such arrangements could not be made, an alternative would be to administer the ASVAB to high school students and select subjects from the tested sample. This approach has several drawbacks. First, the ASVAB takes up to three hours to administer, although the utilization of the seven subtests of interest (Arithmetic Reasoning, Numberican Operations, Paragraph Comprehension, Word Knowledge, Mechanical Comprehension, Automotive Shop Information, and General Science) would cut the time roughly in half. Second, many more students would have to be tested than could be used to compose the four groups described. This is because the extreme scoring individuals are of most interest, and by definition they are scarcer.

The payoff for overcoming the difficulties inherent in either approach would be high. The results would guide the design of training devices best suited to the abilities and talents of the individuals that the Army must train. The results will also provide insight into how training may be better individualized.

Should it prove too costly in time or resources to perform this experiment as described, an alternative would be to proceed as follows. A profile of the average recruit admitted to a representative set of mechanically-oriented career fields could be obtained from the Recruiting Command. This profile could subsequently be used to select subjects. This would result in a less intensive data collection effort, but it has the disadvantage of not addressing the issue of the extremes.

Further Experiments: Different Perceptual-Motor Maintenance Task--The objective of these experiments is to determine if the results of the present experiment generalize to a different perceptual-motor maintenance task. In addition, the task should be more closely related to an actual Army maintenance activity.

Certain maintenance tasks for an internal combustion engine should prove suitable. For example, tuning the engine (gapping spark plugs, adjusting carburetion, etc.) demands coordination of eye, ear, and hand. In addition, objective (electronic) measures of an engine's state of efficiency are

47

readily available through the use of various test instruments. In fact, test instruments along with an engine would be the object of simulation and training.

Experiment 3--The approach to fidelity manipulation in this experiment would be precisely the same as described in Chapter 2 (see Figure 1). Four devices (HH, HL, LH, and LL) would be procured, or designed and built[*] The device specifications would need to be worked out in detail according to the general descriptions in Chapter 2.

The experimental method would parallel the important details of the present experiment. The details of the training (e.g., amount of time) and performance tests would be specified through a series of pilot studies. Particular attention would be given to defining a level of task complexity which will optimize the chances of finding an effect of reduced fidelity.

Experiment 4--This experiment would parallel Experiment 2. The objective will be to determine if intelligence or automotive-shop experience[†] interact with fidelity to an important degree.

Further Experiments: Cognitive Maintenance Task--The objective of experiments in this category is to extend the general approach outlined in Chapter 2 to a cognitive task. Cognitive maintenance tasks are characterized by their non-procedural nature--the technician must troubleshoot in the absence of step-by-step guidance. This entails decisions about which of several alternative actions to take at any given point in time (i.e., to obtain a voltage reading, to remove and replace a component, etc.). Because increasingly complex equipment is more apt to malfunction in unpredictable ways, thus defying a procedural or automatic test solution, soldiers must be able to troubleshoot under conditions of uncertainty.

---

[*]Honeywell has two jeep engines (GFE from the Defense Advanced Research Projects Agency), one of which is functional, that can be used for this experiment.

[†]Automotive-Shop Information is a subtest of the ASVAB.

A major impediment to training research on cognitive tasks has been the lack of objective measures of performance. A recently completed series of studies by Klein and his associates (Klein and Dreyfus, 1982; Klein and Peio, 1982) has resulted in the development of such a measure. These researchers have developed a technique based on predictive accuracy. Basically, subjects are asked to predict moves made in a chess game played by experts. For each successive board position, subjects listed the alternative moves that might have been made and indicated their own choice. Accurate moves were previously defined by Grandmasters who evaluated each move, indicated the number of reasonable alternatives, and judged the quality of each choice. For the most complex moves, proficient players (ratings of 1700 or above) averaged 38% accuracy and novices (ratines of 1300 or below) averaged 23% accuracy, a significant difference. Thus, the utility of this technique was demonstrated for tasks without clear right or wrong answers.

Chess playing is not troubleshooting. Troubleshooting situations, however, are clearly analogous because at each point in time there are a number of reasonable alternative actions, each with a ratable quality. Therefore, it should be possible to adapt the prediction technique to a nonprocedural troubleshooting task. (The technique should also be applicable to tactical decision-making.)

Experiment 5--The objective of this experiment is to determine how reductions in training device fidelity influence training effectiveness for a cognitive maintenance task. The general methodological framework would be identical to the present experiment.

This experiment would be preceded by pilot research to develop a set of test problems. These problems will be related to trouble-shooting of complex aircraft electronics (avionics) using automatic test equipment.[*]

---

[*]Honeywell's Avionics Division builds test equipment and avionics components which would be available for this research.

The basic strategy will be to work with technical experts to construct problems based on actual experience. Each problem would have initial conditions defined by the state of the equipment. Inexperienced and experienced personnel would then be asked to state reasonable action alternatives and select one. The best alternative would be indicated, and the equipment state would be changed in accordance with it. Then a second decision would be made and so on until the malfunction is resolved. Task difficulty or relevant procedural parameters (e.g., time allowed for generating alternatives) would be adjusted until the performance difference between experts and novices is optimized.

Once the validity of this measurement technique is established for nonprocedural troubleshooting, it will be possible to evaluate any differential effects of reduced training device fidelity on task performance. Reductions in fidelity for the selected test equipment and components would be accomplished along the same lines as previously described. The actual training would need to be longer than the five minutes used in the wheel-truing experiment. The results would be analyzed in terms of the proportion of accurate predictions made by the subjects in different conditions. Any differences in performance would be attributable to differences in fidelity.

Experiment 6--The objective of this experiment is to determine the effects of trainee intelligence and electronics aptitude[*] in the context of reduced training device fidelity. When appropriate, the design and procedure would be adapted from Experiment 2.

Organization and Communication of Research Results

The results of the research outlined herein must be accessible to the training device development community. There are several approaches to disseminating the results.

---

[*]Electronics Information is a subtest of the ASVAB.

The widest communication can be achieved by publishing and presenting papers based both on the guiding conceptualization of fidelity and the results of specific experiments. This is being accomplished. R. Hays (in press) will present a review of ARI's overall fidelity research program at the 1982 Interservice Industry Conference in Orlando, Florida. D. Baum (Baum, Riedel and Hays, in press) will present a paper based on the present experiment at the Human Factors Society Annual Meeting in Seattle, Washington.

The results of individual experiments or series of experiments must be integrated into existing and planned guidelines for developing training device specifications, for example, TRADOC Circular 70-82-1. Because the research results will be available on a continuing basis, it is necessary to establish a database that can be easily updated. The most applicable medium is computer mass storage. Hard copies of the database could be published periodically and incorporated as an annex to TRADOC Circular 70-82-1. Also, device developers should be able to gain access to the database from local computer terminals.

How should such a database be organized? We recommend adoption of the general framework presented at the beginning of this chapter. Device developers would specify the task domain, performance context, and task type, and the appropriate subset of the database would be made available. Once in the subset, the search would be organized according to main independent variables—fidelity, task complexity, trainee intelligence, and trainee aptitude. The interaction between the user (device developer) and the database would be guided by menus and prompts. Hard copies of relevant data could be obtained on command.

It is unlikely that the specific research results will be directly relevant to the device developer. Furthermore, there are still bound to be unsanswered questions regarding, for example, how to optimize a device at a particular level of fidelity. In order to overcome these problems, we recommend that a formal training course be developed. This course could be offered through TRADOC and would be mandatory for all personnel assigned to Army Training Device Development. The purpose of the course would be to

51

disseminate state-of-the-art techniques in front-end analysis and the research data necessary to make informed decisions. Familiarization with the fidelity research database would be a key objective of this course.

SUMMARY AND RECOMMENDATIONS

This report presents a conceptual framework to conduct empirical research on the effects of reduced training device fidelity. An experiment was conducted based on the framework.

Bicycle wheel truing was chosen as the experimental task because it is suited to laboratory research, and it is representative of perceptual-motor maintenance tasks. Five devices were procured, or designed and built. The devices varied in physical and functional similarity to the actual equipment. Three devices degraded to the same level both in physical and functional similarity (high, medium, and low). One device was high in physical and low in functional similarity, and the final device was low in physical and high in functional similarity. Twenty naive high school and vocational technical school students were trained on each device before performing on the actual equipment.

The results indicated all devices resulted in learning; however, there was not a significant effect of devices differing in overall level of fidelity. Training with low fidelity line drawings resulted in performance not significantly lower than training on the actual equipment. However, examining the separate effects of physical and functional similarity revealed that physical similarity was significant. Functional similarity did not achieve significance. For training perceptual-motor maintenance tasks, this experiment indicates that high physical similarity is important, but high functional similarity adds no further performance benefit.

The definition of fidelity that guided this research was discussed and several refinements of the definition were indicated including an elaboration of the definition of physical similarity and an extension of the level of abstraction for functional similarity.

The research proposal presented in this chapter focuses on individually performed perceptual-motor and cognitive maintenance tasks, but a general framework for further research was also presented. If carried out, the recommended research will replicate and generalize the present results and examine the effects of trainee intelligence and aptitude. Further research is recommended on a more complex wheel-truing task and on a perceptual-motor maintenance task involving an internal combustion engine. Also, research on a cognitive maintenance task, nonprocedural troubleshooting of aircraft electronics, is recommended.

Finally, means for organizing and disseminating the research results were discussed. It was recommended that a computer database be created to direct future fidelity research and to provide a basis for delivering guidance to training device developers.

References

Baum, D.R., Riedel, S.L., & Hays, R.T. Training device fidelity and transfer of a perceptual-motor skill. To appear in Proceedings of the Human Factors Society 1982 Annual Meeting. Seattle, WA, in press.

Baum, D.R., Smith, D.A., Hirshfeld, S.F., Klein, G.A., Swezey, R.W., & Hays, R.T. Specification of training simulator fidelity: A research plan. Alexandria, VA: US Army Research Institute, Technical Report 558, February, 1982. (AD A127 346)

Hays, R.T. Simuator fidelity: A concept paper. Alexandria, VA: US Army Research Institute, Technical Report 490, November, 1980. (AD A101 475)

Hays, R.T. A fidelity definition and outline of interactive factors. R.T. Hays (Ed.), Research Issues in the Determination of Simulator Fidelity: Proceedings of the ARI sponsored workshop, 23-24 July 1981, Alexandria, VA: US Army Research Institute, Technical Report 547, November, 1981.

Hays, R.T. (Ed.) Research issues in the determination of simulator fidelity: Proceedings of the ARI sponsored workshop, 23-24 July 1981. Alexandria, VA: US Army Research Institute, Technical Report 547, November, 1981. (AD A118 253)

Hays, R.T. ARI's research program to determine training simulator characteristics. To appear in Proceedings of the Third Annual Interservice Industry Training Conference. Orlando, FL: Naval Training Equipment Center, in preparation.

Johnson, W.B. Computer simulators in fault diagnosis training: An empirical study of learning transfer. Champaign, IL: Coordinated Science Laboratory Report T-101, University of Illinois, 1980.

Johnson, W.B., & Rouse, W.B. Computer simulations for fault diagnosis training: From simulation to live system performance. Paper presented at the Annual Meeting of the Human Factors Society, Los Angeles, CA, October 1980.

Keppel, G. Design and Analysis: A Researcher's Handbook. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1973.

Klein, G.A., & Dreyfus, S. Recognitional vs calculational capacities in highly proficient performance. In preparation, 1982.

Klein, G.A., & Peio, K.J. The use of a prediction paradigm to study proficient performance. In preparation, 1982.

McCormick, E.J. Human Factors Engineering. New York, NY: McGraw-Hill, Co., 1970.

PREVIOUS PAGE
IS BLANK

Mirabella, A. ARI Simulation Research Program. In R.T. Hays (Ed.),
     Research Issues in the determination of Simulator Fidelity:
     Proceedings of the ARI sponsored workshop, 23-24 July 1981, Alexandria,
     VA: US Army Research Institute, Technical Report 547, November 1981.

Office of the Assistant Secretary of Defense (MRA&L). Profile of American
     Youth: 1980 nationwide administration of the Armed Services Vocational
     Aptitude Battery. Washington, DC: OSD (MRA&L), 1982.

Orlansky, J. & String, J. Cost-effectiveness of maintenance simulators for
     military training. Arlington, VA: Institute for Defense Analyses, IDA
     Paper P-1568, August 1981.

Rouse, W.G. Problem solving performance of maintenance trainees in a fault
     diagnosis task. Human Factors, 21, pp. 195-203, 1979.

Williams, M., Hollan, J., & Stevens, A. An overview of STEAMER: An advanced
     computer-assisted instruction system for propulsion engineering.
     Behavior Research Methods & Instrumentation, 13, pp. 85-90, 1981.